Snap and Translate Using Windows Phone

Jun Du, Qiang Huo, Lei Sun, Jian Sun
Microsoft Research Asia, Beijing, P. R. China
Emails: {jundu, qianghuo, v-lesun, jiansun}@microsoft.com

I. Abstract

For a language learner or a tourist traveling in a foreign country, OCR translation using a Smartphone with a camera will allow the person to read a menu/sign in a foreign language very conveniently. This has attracted much research activities in the past decade. Given the good quality of the built-in camera and the enough computational capability of Smartphones on the market, several commercial mobile OCR translation apps have been released, which can be divided into two broad classes: client only app and client-plus-cloud app. For the former type of apps, all the processing is done on the smartphone itself. Pleco and Word Lens are two popular examples. Pleco is a Chinese Dictionary app which is designed for language learning. Word Lens is an iPhone app which can translate a sign between English and Spanish with a live augmented reality (AR) overlay of the translation result. In both cases, only word by word translation (or dictionary lookup) is performed. For client-plus-cloud apps, memory and computation demanding tasks such as OCR and translation are done typically in the cloud, while other functions are run on the client-side. By definition, network access should always be available to enable a client-plus-cloud implementation. The translation feature in Google Goggles is such a representative, which allows a user to translate sentence(s) by taking a picture and drawing a precise bounding box of the intended text, because more advanced OCR and translation technologies can be implemented in the cloud.

Recently, we have also developed a prototype of a mobile app called “Snap and Translate” on “Windows Phone 7” (WP7) based on a client-plus-cloud architecture. In one of the operation modes, a user can use one of three natural gestures, namely *tap* a word or *swipe* a phrase or *circle* a paragraph with a finger on the captured text image to indicate his/her intention explicitly. The intended text image patch will then be extracted automatically on the phone and sent to the cloud for recognition and translation. This makes our system different from the aforementioned apps with the following benefits: 1) a more accurate text extraction with the help of user’s intention, 2) less computations on the phone without processing the whole image, therefore longer battery time, and 3) less network traffic and smaller latency because only a small image patch is sent over the network.

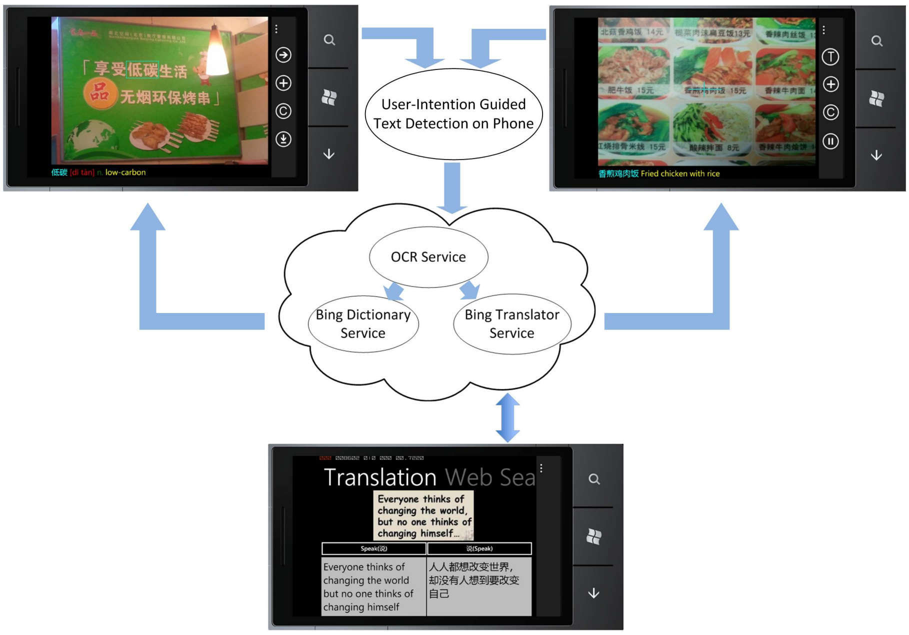
「享受低碳生活
品 无烟环保烤串」
低碳 [dī tàn] n. low-carbon
User-Intention Guided Text Detection on Phone
北菇香鸡饭 14元
榄菜肉沫扁豆饭13元
香辣肉丝饭
肥牛饭 15元
香煎鸡肉饭 15元
香辣牛肉面
红烧排骨米线 15元
酸辣拌面 8元
香辣牛肉烩饼
香煎鸡肉饭 Fried chicken with rice
OCR Service
Bing Dictionary Service
Bing Translator Service
Translation Web Sea
Everyone thinks of changing the world, but no one thinks of changing himself...
Speak(说)
说(Speak)
Everyone thinks of changing the world but no one thinks of changing himself
人人都想改变世界，却没有人想到要改变自己

Fig. 1. Overall system architecture.